ENTRIX
ENVIRONMENTAL CONSULTANTS

A Review of Proposed TPH Sediment Quality Values and an Alternative Method to Define Hydrocarbon Values for Portland Harbor

1. PRIMARY ISSUE

Organic compounds quantified as "Total Petroleum Hydrocarbons" (TPH) are ubiquitous in the Lower Willamette (LW) River sediments, reflecting contributions from a wide variety of compounds of differing levels of toxicity from disparate sources. It has been proposed that TPH values be used as a screening tool for sediment toxicity throughout the LW Site. While setting a TPH Sediment Quality Value (SQV) may be tractable for compositionally-similar (preferably identical) TPH mixtures in different samples, pursuing this approach for compositionally different samples on a river-wide basis is an impossible task for a number of reasons described in this document.

At present, there is no scientifically-defensible method to equate the toxicity of samples using TPH measurements, because the samples are compositionally dissimilar from a TPH perspective. Hence, there is no defensible method for the use of TPH measurements as SQVs as proposed in the screening assessment of the LW sediments.

Nevertheless, it is recognized that the LW Group (LWG) faces some constraints in recommending whether and how TPH measurements will be used in Round 2 in helping to define initial Areas of Potential Concern (AOPCs). These practical limitations were discussed in a meeting with the common LWG consultants on September 8, 2006. However, there was general acknowledgment of the overall inadequacies of TPH measures for use in developing reliable SQVs, and of the need to reflect those inadequacies clearly in the Round 2 report. In particular, to improve the defensibility of any hydrocarbon SQV, it will be essential that the Round 2 report clearly distinguish applicable hydrocarbon distributions, discuss TPH issues in the uncertainties section of the report, and identify appropriate ways to resolve those issues as data gaps to be filled in Round 3 as needed with respect to specific AOPCs.

2. BACKGROUND

As part of the Phase 2 RI process it is necessary for the LWG working with their common consultants to propose SQVs for a range of contaminants. The use of TPH as a measure of toxicity has been proposed for use directly as a measure of toxicity as well as in lieu of using PAH concentrations. However, the TPH SQVs proposed for this purpose in the Round 2 submittal are fatally flawed for a number of scientific reasons, and cannot serve their intended purpose; hence rather than using TPH measurements to develop SQVs, it makes more sense and is more scientifically defensible to use PAH data (for which toxicity data are available).

As tested for by Columbia Analytical Services in the Round 2A Portland Harbor study, TPH covers hydrocarbons of any description and many other organic compounds that could be measured analytically within the carbon range of *n*-C_{12} to *n*-C_{40}, regardless of source. This range excludes volatile hydrocarbons such as those most characteristic of

gasoline (e.g., benzene), but includes hundreds of organic compounds (hydrocarbons and non-hydrocarbons) that make up materials such as kerosenes; jet fuel; mineral oil; vegetable oils; diesel and bunker oils; soot from industrial, commercial, residential and natural combustion processes; used motor oil; slop oil; asphalt and tars; coal (anthracite, bitumen, lignite); peat; creosote; plant waxes and essential oils; and the microbial breakdown products of both plant and animal biomolecules. These hydrocarbons are often categorized as biogenic (from a natural source such as plant material), petrogenic (from a fossil fuel source such as oil or coal) or pyrogenic (from a natural or man-made combustion source, including those creating soot or used to produce creosote from coal tar or wood).

Diesel Range Hydrocarbon (DRH) is a measure of an analytically-defined mixture of organic compounds that, for the Round 2A study, comprise the TPH from *n*-C_{12} to *n*-C_{25}. Residual Range Hydrocarbon (RRH) is a measure of the hydrocarbons that comprise the remaining TPH from *n*-C_{25} to *n*-C_{40}. These terms do not reflect the actual constituents (identified above) in each sample, but are used simply because fuels (typically diesel fuel #2 and motor oil) are used as laboratory standards to convert the analytical instrument signal in the sample to reported DRH and RRH concentration in each sample. Hence, three samples, each containing 100 mg/kg (parts per million) DRH, could contain (1) 100 mg/kg diesel fuel; (2) 100 mg/kg of the natural (biogenic) hydrocarbons from tree leaves within the DRH range; and (3) 100 mg/kg of the hydrocarbons from soot in highway runoff within the DRH range. Alternatively, the three samples could contain varying proportions of a degraded mixture of hydrocarbons from all three sources. The toxicity of hydrocarbons from these three sources varies substantially and so "100 mg/kg DRH" cannot be meaningfully related to a toxic effect without knowledge of the source(s), and more specifically, of the particular hydrocarbon composition that comprises the 100 mg/kg DRH.

In order to derive a potential toxicity value that could be used to screen sediments for further investigation, samples using the DRH and RRH designation were used with laboratory toxicity test results to determine concentrations of hydrocarbon mixtures at which a potential for adverse biological effects would be predicted (i.e., SQVs). That process requires that a given concentration of DRH or RRH represents the same mixture of constituents and toxicity across samples. For some specific compounds defined analytically, i.e., chlorinated dioxin and furan congeners (e.g., 2,3,7,8-tetrachlorodibenzodioxin), metals (arsenic, copper), or other organic compounds (lindane, DDT isomers, PCB congeners), specific concentration vs. toxicity criteria have been established over years of peer-reviewed research. In contrast, the relationship between DRH and RRH to toxic thresholds has not been defined due to the myriad different compounds that are included within these "terms of art."

In sum, the central difficulty with the proposed SQVs for DRH and RRH is that no dose-response relationship can be discerned where samples of hydrocarbon mixtures have an unknown, variable composition of constituents with variable toxicities. Therefore, the DRH and RRH measures, particularly when not controlled for changes in composition among samples, are not comparable to single chemical or chemical mixture concentrations that have been accepted in the scientific literature for use in SQV development. The causative agents are unknown for the observed correlation between

DRH and some toxicity measures, as well as for RRH and some toxicity measures. In fact, the causative agents may differ among areas sampled for TPH, resulting in apparent toxicity correlations that simply result from the broad distribution of hydrocarbons in areas that contain elevated levels of other contaminants from varying sources. Therefore, TPH measures should not be used as predictors of adverse biological effects in other areas without understanding what an exceedance of a DRH or RRH SQV represents.

Consequently, the LWG approach raises the legitimate concern that given the lack of a scientific basis on which to specify SQVs for DRH and RRH, any value proposed as a SQV would be arbitrary. However, we understand that such a value is being demanded by the EPA for use at this current stage of the Ecological Risk Assessment (ERA). In addition, the September 8, 2006 meeting with the common LWG consultants made clear that the LWG had limited options to resolve scientific concerns before identifying AOPCs in the Round 2 report. While recognizing the shortcomings of the present approach for using TPH measurements to define SQVs, we also propose an approach to define "temporary hydrocarbon SQVs" that would be appropriate for use until research is completed to justify (or not) a TPH toxicological endpoint. These alternative SQVs will be useful in the Round 2 report to illustrate the uncertainty of the existing proposed SQVs, particularly the 340 mg/kg SQV for DRH.

3. CRITICAL FLAWS WITH THE DRH/RRH APPROACH

The methodology EPA requested Windward et al. (2006) to use for developing the screening levels proposed for DRH and RRH did not account for the sample heterogeneity in DRH and RRH measurements. As such, the derived SQVs from these bulk hydrocarbon concentrations are meaningless for the intended purpose of screening the potential toxicity of sediments in subsequent risk assessment. *While it is recognized that petroleum-derived hydrocarbons are of concern to the EPA, it is critical that regulatory attention focuses specifically on such compounds, rather than on an arbitrary selection of hydrocarbons that simply co-occur with but are irrelevant to petroleum-specific releases*.

For example, the 340 mg/kg SQV proposed for DRH for both Level 2 ("minor effects") and Level 3 ("moderate effects") is of greatest concern because that value resulted from the use of flawed assumptions. Specifically, screening of sediments using the proposed DRH SQV is projected to ***incorrectly*** identify as Level 3 approximately **12%** of 395 samples above the median DRH concentration in all Portland Harbor studies, approximately **13%** of the 142 Round 2a samples for which there are DRH results, and **51%** of Round 2a sediments above the proposed DRH SQV. That is, there was only a 50:50 chance that a sediment sample above the proposed DRH SQV was actually found toxic in the Round 2a study. Hence using this SQV in a predictive screening mode with the 340 mg/kg cutoff presents a high degree of scientific inaccuracy, with a 50% false positive error rate that greatly exceeds the study goal of a maximum 20%. If the anomalously large GASCO-related hits are removed there is an even larger discrepancy (approximately 58%). In contrast, proposed SQVs for other COPCs have a reasonable toxicological basis and appear to be predictive of adverse effects in Portland Harbor. The concern is not associated with use of empirical models such as the Floating Percentile Model (FPM) to develop SQVs, but the application to single-value TPH measures.

The 2,700 mg/kg SQV for RRH at Level 2 and the 4,500 mg/kg SQV for RRH at Level 3 were also derived by the same approach, hence suffer from similar flaws because they are single-values representing a mixture of unknown and varying composition.

The critical concerns with the proposed numerical DRH and RRH SQVs are:

- The justification offered for use of the DRH and RRH SQVs rests on their consistency with a narcosis model of petroleum-based toxicity caused by a wide range of hydrocarbons. A more complete characterization of the literature cited in the report and additional peer-reviewed literature[1] illustrates that there is no "similar" narcotic effect occurring across all hydrocarbon groups but rather an effect that is essentially solubility-dependent. In other words, the narcotic "potential" is directly proportional to the amount of each hydrocarbon that can be dissolved from sediments, not to the total hydrocarbon concentration measured in sediments (DiToro et al. 2000). The most extreme acute narcotic effects would be expected in samples containing high concentrations of relatively soluble hydrocarbons. Hence, narcosis as the claimed toxicological basis for use of the SQVs is suspect for four reasons:

    1. The most acutely toxic (narcotic) hydrocarbons are the most soluble and volatile compounds which are also least likely to persist in sediments, especially after transport away from their release point (i.e., they are the most subject to degradation, or "weathering" processes). However, the chromatograms associated with the samples used to establish the DRH and RRH SQVs show that such acutely toxic hydrocarbons are either absent or disproportionately low in samples with a potential petrogenic contribution. Those hydrocarbons are even disproportionately low in the rare sample that might contain a minor diesel component;
    2. Fuels with hydrocarbons that have molecular weights higher than a particular solubility "cut-off" show little or no measurable toxicity (CONCAWE 2001). This cut-off varies predictably with the chemical structure of the hydrocarbon and is consistent with the Veith et al. (1983) paper cited in the Windward et al (2006) report. For example, the solubility cut-off for *n*-alkanes is estimated to be approximately *n*-C_{10} but for alkyl benzenes is about C_{14} (CONCAWE 2001). Narcosis as a contributor to fish embryo mortality has been associated with naphthalenes (Barron et al. 2004), which are relatively volatile $C_{10} - C_{14}$ aromatic compounds, depending on the extent of alkylation.
    3. It is not surprising that a derived SQV for RRH would be higher than that for DRH based on the inert characteristics of most hydrocarbons that comprise RRH (CONCAWE 1998). However, whether there is an appropriate RRH SQV for Portland Harbor sediments, or whether there is even an apparent correlation of RRH with toxicity, is uncertain. Hydrocarbons that comprise RRH are nearly or totally insoluble under environmental conditions, indicating that if narcosis is the valid toxic mechanism, the samples should

[1] Connell and Markwell 1992; Veith et al. 1988; Barron et al. 2004; CONCAWE 2001 and cited references; DiToro and McGrath 2000; CONCAWE 1998 and cited references; van Wezel and Opperhuizen, 1995.

have relatively little to no toxicity compared to that potentially associated with hydrocarbons that comprise DRH. The *n*-C_{25} to *n*-C_{40} carbon range for RRH analyses far exceeds the solubility "cut-off" of *n*-C_{10} for alkanes and C_{14} "cut-off" for alkylbenzenes (CONCAWE 2001) where narcotic toxicity has been reasonably associated with hydrocarbons.

4. The apparent correlations identified by Windward et al. (2006) should be further examined after separating out the confounding factors of mixed source contributions to "DRH" and "RRH". Explicitly, DRH and RRH concentrations could have an observed correlation with toxicity, not because they are the best predictors of toxicity, but because they are correlated with the best predictors. That concern is heightened by the high colinearity and covariance of DRH and several other benthic COPCs such as PAHs, ammonia and sulfides, and the sensitivity of the FPM to the order in which the COPCs are added to the model. Some relatively simple diagnostics, such as looking for outliers and influential data points in residual plots or basic statistical methods (e.g., a residual test using the t-test distribution) could be used to test this hypothesis. In any event, a claim that TPH measures represent the majority of the toxicity associated with petroleum compounds in empirical models used to develop SQVs is unwarranted because that observation has not been ground-truthed to determine if "TPH", "DRH", or "RRH" are comparable or meaningful measurements at the Round 2 sampling locations.

- The relationship between observed toxicity and hydrocarbon compositions is not consistent among samples (Table 1; Figures 1-5), more likely than not due to the variable toxicity of the constituents of the mixtures in samples exhibiting similar identical DRH or RRH concentrations. Therefore, there is no toxicological basis for using DRH and RRH concentrations to develop SQVs (i.e., that there is no showing of a causative relationship between narcosis and observed DRH and RRH distributions, that would result in statistically significant narcotic effects at concentrations above the SQVs). If the differential toxicity of hydrocarbons comprising DRH and RRH was not, and cannot be, accounted for in the single DRH and RRH concentration values, their proposed use as SQVs represents the use of essentially arbitrary numbers as measures of toxicity.

- Non-petrogenic sources account for the dominant hydrocarbon component in many samples in the LWG database, with those samples widely distributed throughout the Site. In particular, many samples in the concentration range of the proposed 340 mg/kg SQV for DRH appear primarily to contain contributions of natural and typical "urban-derived" hydrocarbons with the following characteristics:

    1. A consistently higher boiling range of hydrocarbons than found in diesel oil.
    2. A dominant pattern of natural biogenic and pyrogenic hydrocarbons, with a minor and substantially weathered petrogenic component characteristic of common land-based runoff sources and atmospheric fallout.
    3. A substantial combustion PAH component in all samples.

High concentration samples almost exclusively represent localized creosote/tar sources, which are not comparable in toxicity with the potential toxicity of petrogenic sources or of typical urban sources of hydrocarbons. In fact, if the samples had been properly categorized by source type, it would substantially weaken the claimed "stronger" correlation between DRH and RRH concentrations and toxicity results versus correlations for other chemical measures.

- The proposed petroleum-based mixture SQVs are markedly less reliable than other SQV measures used in the screening model, producing far more false positives than the SQV for any other constituent (18 false positives at Level 3 for DRH versus none or one[2] for other SQVs).

4. AN ALTERNATIVE PROPOSAL

As the above discussion demonstrates, the proposed petroleum-based mixture SQVs lack both the toxicity correlation and toxicological basis required to establish a toxicity relationship. Absent these bases, the proposed use of the 340 mg/kg DRH, which minimizes false negatives but results in numerous false positive predictions, represents merely expedience, not good science. An improvement would be to ensure that the composition of hydrocarbons in the petroleum-based mixtures varies as little as possible among samples used to develop the SQVs, and then develop an SQV specific to that particular hydrocarbon composition as described below.

Alternatively, if the existing DRH and RRH SQVs must be used for screening to help identify preliminary AOPCs in the Round 2 report, then a reasonable solution would be to use SQVs with no associated false positives. A DRH SQV of 4,200 mg/kg and the RRH Level 3 SQV of 4,500 mg/kg ("temporary hydrocarbon SQVs") both have that characteristic. Such a SQV could be used in uncertainty discussions within the Round 2 report, to identify data gaps, and to contrast the extent of AOPCs using the proposed SQVs and alternative SQVs identified in this section.

The samples from which these alternative SQVs are derived resemble urban-derived hydrocarbon mixtures and, therefore, these higher SQVs should still only be applied to screen samples dominated by that fingerprint, which is likely to include much of the river channel and the areas affected by stormwater runoff. However, the fingerprint of TPH chromatograms should be examined to ensure that it is appropriate to apply these screening SQVs.

A petrogenic hydrocarbon SQV or creosote/tar source SQV would have different characteristics than the proposed "urban-derived" hydrocarbon SQVs and the development of such SQVs require further assessment. Based on a review of Portland Harbor data, petrogenic or creosote/tar sources dominate in a relatively small number of localized areas near obvious sources. Therefore, the immediate need for screening is lessened.

---

[2] At Level 2, the DRH SQV produced 15 false positives while SQVs other than ammonia and sulfides produced none or one. Ammonia and sulfides still only produced 7 and 9 false positives, respectively, substantially fewer than the DRH SQV.

The tradeoff in using the DRH SQV at a level with no associated false positives is an increase in the overall "false negatives"[3] resulting from using the proposed SQVs for the other COPCs. Based on the review to date, there is no reason to believe that DRH compounds are necessarily the causative agent, or correlated with the causative agent, of toxicity at the locations that would no longer be "captured" by the DRH SQV. To avoid speculation, we recommend the following:

1. Reassess the specific samples at "false negative locations" to determine if there is a basis for associating a particular chemical or chemical class with the observed toxicity, even if those chemicals have been previously eliminated from SQV development.
2. Reassess whether the SQV study erred in using total PAH as the sole measure for that chemical class and the effect that error may have had on false negatives.[4]

In summary, it is important that the SQVs capture the essence of petroleum releases rather than representing arbitrary "background noise." A major data gap left from the Round 2 study is how to properly characterize chemical-biological relationships for hydrocarbon mixtures. To address that issue, as needed on an AOPC-specific basis, additional field studies could be scoped and completed in a timely manner to determine if a scientifically-defensible DRH and RRH SQV can be developed for hydrocarbon-specific sources types.

---

[3] Increasing the DRH SQV from 340 mg/kg to 4,200 mg/kg eliminates the 18 false positives at Level 3, the 15 false positives at Level 2, and the 13 false positives at Level 1. The higher SQV increases the number of false negatives (samples in which the toxicity is unexplained by any of the SQVs) by seven at Level 2 and by eight at Level 3 Based on the Round 2a data set for which there are hydrocarbon and toxicity results, we estimate that the increased number of false negative samples increases the percent false negatives to approximately 27% at Level 2 and to approximately 39% at Level 3, assuming that the acute mortality exhibited by samples coded "-1" in the database should appropriately identify conditions that are acutely toxic for growth. Although these percentages also exceed the goal of <20% false negatives the samples with "unexplained" toxicity must not be incorrectly attributed to a COPC without a suitable scientific basis. Of these additional "unexplained" samples, one appears to have been incorrectly reported as containing elevated concentrations of diesel-range hydrocarbons and should have been reported as containing almost exclusively residual-range hydrocarbons. In fact, three of the six samples are dominated by residual-range hydrocarbons, not diesel-range hydrocarbons. The remaining samples contain a mixture of degraded diesel-range hydrocarbons and residual-range hydrocarbons. Whether or not the DRH is an appropriate surrogate for toxicity in these samples has not been adequately studied.

[4] There was no reasonable justification for summing low molecular weight PAH (LPAH) compounds with high molecular weight PAH (HPAH) compounds to develop a total PAH SQV, because PAH compounds within these two groups are dominated by different sources with different bioavailabities. The LWG statistical analysis found a strong correlation among HPAH and a somewhat lesser correlation among LPAH, consistent with source and environmental fate considerations (which produce more variability in LPAH distributions). A different approach for developing PAH SQVs and a better classification of TPH might have resulted in fewer false negatives to the extent that toxicity is attributable to hydrocarbons, or hydrocarbons are correlated with a toxic agent.

References

Barron, M.G. M.G. Carls, R. Heintz, and S.D. Rice. 2004. Evaluation of fish early life-stage toxicity models of chronic embryonic exposures to complex polycyclic aromatic hydrocarbon mixtures. *Toxicol. Sci.* 78:60–67.

Connell D and R.D. Markwell. 1992. Mechanism and prediction of nonspecific toxicity to fish using bioconcentration characteristics. *Ecotoxicol. Environ. Saf.* 24:247-265.

CONCAWE. 2001. Environmental classification of petroleum substances - summary data and rationale. Report No. 01/54. Prepared by the Petroleum Products Ecology Group. CONCAWE. Brussels. 144 pp.

CONCAWE. 1998. Heavy fuel oils. Product dossier No. 98/109. Prepared by the Petroleum Products and Health Management Groups. CONCAWE. Brussels. 48 pp.

DiToro D.M. and J.A. McGrath. 2000. Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria. II. Mixtures and sediments. *Environ. Toxicol. Chem.* 19(8):1971–1982.

van Wezel, A.P. and A. Opperhuizen. 1995. Narcosis due to environmental pollutants in aquatic organisms: residue-based toxicity, mechanisms, and membrane burdens. *Crit. Rev.Toxicol.* 25:255-279.

Veith, G.D., D.J. Call, and L.T. Brooke. 1983. Structure-toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci.* 40:743-748.

Windward Environmental, Avocet Consulting, and TerraStat Consulting Group. (2006) **DRAFT** Portland Harbor Superfund Site Ecological Risk Assessment: Interpretive Report: Estimating Risks to Benthic Organisms Using Predictive Models Based on Sediment Toxicity Tests. Report WE 06-002. Prepared for the Lower Willamette Group. March 17, 2006. 78 pp. *LWG Report Do Not Cite*.